

## Data Import and Transformation



**Data wrangling** is the process of cleaning and transforming data to make it more appropriate for data analysis. The process generally follows these main steps:

- Explore the raw data and check the general quality of the dataset.
- Transform the raw data, by restructuring, normalizing, and cleaning the data. For example, this could involve handling missing values and detecting errors.
- Validate and publish the data.

Data wrangling is an *iterative* process where you do some data transformation then check the results and come back to the process to make improvements.

QUIZ QUESTION

See if you can match the following descriptions with the data wrangling task they describe.

*Submit to check your answer choices!*

DESCRIPTION	TASK
Counting the number of missing values	Data discovery and exploration
Missing values imputation	Data cleansing
Normalize feature values	Data restructuring

SUBMIT

The best way to understanding data wrangling is to get hands-on experience actually doing it—so that's what we'll do in our next lab.

